

---

# Phase Transitions, Distance Functions, and Implicit Neural Representations

---

Yaron Lipman<sup>1 2</sup>

## Abstract

Representing surfaces as zero level sets of neural networks recently emerged as a powerful modeling paradigm, named Implicit Neural Representations (INRs), serving numerous downstream applications in geometric deep learning and 3D vision. Training INRs previously required choosing between occupancy and distance function representation and different losses with unknown limit behavior and/or bias. In this paper we draw inspiration from the theory of phase transitions of liquids and suggest a loss for training INRs that learns a density function that converges to a proper occupancy function, while its log transform converges to a distance function. Furthermore, we analyze the limit minimizer of this loss showing it satisfies the reconstruction constraints and has minimal surface perimeter, a desirable inductive bias for surface reconstruction. Training INRs with this new loss leads to state-of-the-art reconstructions on a standard benchmark.

## 1. Introduction

*Implicit neural representation* (INR) refer to representing a surface in  $\mathbb{R}^3$ , or more generally a hyper-surface in  $\mathbb{R}^d$ , as a zero level-set of a neural network  $f$  with parameters  $\theta$ :

$$\mathcal{S}_\theta = \{ \mathbf{x} \in \mathbb{R}^d \mid f(\mathbf{x}; \theta) = 0 \}. \quad (1)$$

Recent works have demonstrated the benefits in this representation (Mescheder et al., 2019; Chen & Zhang, 2019; Park et al., 2019; Atzmon et al., 2019), including: the flexibility and expressive power of neural networks  $f$ , their favorable inductive bias / implicit regularization properties, and the facilitation of downstream learning applications involving 3D shapes. Indeed, implicit neural representations find more and more applications in machine learning, computer vision, and computer graphics (Saito et al., 2019; Niemeyer et al., 2020; 2019; Gropp et al., 2020; Jiang et al.,

---

<sup>1</sup>Facebook AI Research <sup>2</sup>Weizmann Institute of Science. Correspondence to: Yaron Lipman <ylipman@fb.com, yaron.lipman@weizmann.ac.il>.

Preprint.



Figure 1. Implicit neural representations learned with the novel PHASE loss. This new loss unifies occupancy and signed distance representations, and minimizes the perimeter of the reconstructed surface.

2020; Sitzmann et al., 2020; Yariv et al., 2020; Tancik et al., 2020).

Still, the problem of *training* an implicit neural representation from raw data  $\mathcal{X} \subset \mathbb{R}^d$  such as a point-cloud or a *triangle soup*<sup>1</sup> is not fully solved.

Some previous works reduce the problem to a supervised regression problem (Park et al., 2019; Mescheder et al., 2019; Chen & Zhang, 2019). That is, first precomputing (or receiving as input) an implicit representation to the geometry  $\hat{f}$  and in a second step optimizing the network  $f$  to approximate  $\hat{f}$ . Other works consider the unsupervised version where only the raw input geometry is provided as input  $\mathcal{X} \subset \mathbb{R}^d$  (Atzmon & Lipman, 2020a; Gropp et al., 2020; Sitzmann et al., 2020; Tancik et al., 2020; Williams et al., 2020). Two limitations of previous approaches are: First, the limit behavior and/or inductive bias of their loss/representation is often not clear, sometimes leading to undesired artifacts in the reconstructed geometry. Second, they require choosing between different implicit representations such as Signed Distance Functions (SDF) or occupancy functions.

The goal of this paper is to introduce a new loss for training INRs from raw data that unifies the different implicit representations and that its limit behavior is theoretically

---

<sup>1</sup>A *triangle soup* is a collection of triangles, not necessarily consistently oriented or defining a 2-manifold.

well understood. Motivated by phase transition theory of liquids we design a one parameter ( $\epsilon$ ) loss that learns a signed density  $u$ , that converges (as  $\epsilon \downarrow 0$ ) to a proper occupancy function, where its log-transform  $w$  approximates the signed distance function. Furthermore, we prove that the zero level-set of the minimizers of this loss converge to a limit surface  $\mathcal{S}$  that is guaranteed to pass in vicinity of the input data  $\mathcal{X}$ , and has minimal surface perimeter. These properties serve as a good inductive bias for the surface reconstruction problem, and we show that in practice this loss achieves superior reconstructions compared to relevant baseline methods. Furthermore, its minimal perimeter property makes it more suitable for high frequency INR training methods, such as (Tancik et al., 2020; Sitzmann et al., 2020), considerably reducing the ghost artifacts created when training with less structured losses. Figure 1 shows an example of two INR surfaces learned with this novel loss (PHASE).

To summarize our contributions, we introduce a novel loss for training INRs directly from input raw data that is supported by a well established limit theory and state-of-the-art reconstruction results. In the limit, the minimizers  $u_0$  of the PHASE loss satisfy the following desirable properties: (i) *Proper occupancy*:  $u_0 \leq 0$  defines a finite perimeter boundary surface  $\mathcal{S}$ ; (ii) *Zero reconstruction error*: The surface  $\mathcal{S}$  passes near the input data  $\mathcal{X}$ ; (iii) *Minimal perimeter*:  $\mathcal{S}$  has minimal surface perimeter; and (iv) *Signed distance function*: The log transform of  $u_0$  approximates the signed distance function to  $\mathcal{S}$ .

## 2. Preliminaries

In this section we provide some preliminaries and background for the main constructions and proofs in the next section. We work in  $\mathbb{R}^d$ , and  $\|\mathbf{x}\|$  will denote the standard euclidean norm of a vector  $\mathbf{x} \in \mathbb{R}^d$ . A subset  $\Omega \subset \mathbb{R}^d$  is called a *Lipschitz domain* if it is open, bounded, connected, and has a boundary that is the union of finitely many Lipschitz manifolds;  $|\Omega| = \int_{\Omega} 1$  is the total volume of  $\Omega$ .  $L^p(\Omega)$  is the space of scalar functions,  $f : \Omega \rightarrow \mathbb{R}$ , with finite Lebesgue integral  $\int_{\Omega} |f|^p < \infty$ ; and  $C_c^\infty(\Omega)$  denotes the space of infinitely smooth functions with a compact support in  $\Omega$ .  $L^p(\Omega; \mathbb{R}^d)$ ,  $C_c^\infty(\Omega; \mathbb{R}^d)$  will denote vector valued functions  $\mathbf{v} : \Omega \rightarrow \mathbb{R}^d$  with each coordinate in  $L^p(\Omega)$ ,  $C_c^\infty(\Omega)$ , respectively.

### 2.1. Sobolev spaces

Sobolev spaces generalize the classical notion of differentiable functions to *weakly* differentiable functions. They are useful in the field of partial differential equations and calculus of variations since, in contrast to classical differentiable functions, they survive the process of limit.  $W^{1,p}(\Omega)$ ,  $p \in [1, \infty]$ , is the Sobolev space of all  $L^p(\Omega)$  functions with first *weak derivatives* also in  $L^p(\Omega)$  (definition of weak

derivatives and more details on Sobolev spaces is provided in (Adams & Fournier, 2003; Gilbarg & Trudinger, 2015; Leoni, 2017)). One interesting relation to neural networks is the following lemma proved in the supplementary:

**Lemma 1.** *Neural networks with ReLU activations are in  $W^{1,p}(\Omega)$ , for all  $p \in [1, \infty]$ , if  $\Omega \subset \mathbb{R}^d$  is a Lipschitz domain.*

### 2.2. Functions of bounded variation and perimeter

The space of bounded variation functions over  $\Omega$ , denoted  $BV(\Omega)$ , is defined as the collection of all functions  $f \in L^1(\Omega)$  such that

$$\int_{\Omega} \|\nabla f\| \stackrel{\text{def}}{=} \sup_{\substack{\mathbf{v} \in C_c^\infty(\Omega; \mathbb{R}^d) \\ \|\mathbf{v}\| \leq 1}} \int_{\Omega} f \operatorname{div}(\mathbf{v}) < \infty \quad (2)$$

where  $\|\mathbf{v}\| \leq 1$  means that for all  $\mathbf{x} \in \Omega$ ,  $\|\mathbf{v}(\mathbf{x})\| \leq 1$ , and  $\operatorname{div}$  is the divergence operator.  $BV(\Omega; \mathcal{A})$  is the set of bounded variation functions over  $\Omega$  with image in the set  $\mathcal{A}$ .

*Surface perimeter* is a generalization of surface area to boundaries of general sets: Given a subset  $\mathcal{I} \subset \Omega$ , its perimeter in  $\Omega$  is defined by

$$\operatorname{per}_{\Omega}(\mathcal{I}) = \int_{\Omega} \|\nabla \mathbf{1}_{\mathcal{I}}\| \quad (3)$$

where  $\mathbf{1}_{\mathcal{I}}$  is the indicator function of the set  $\mathcal{I}$ . If  $\mathcal{I}$  has a smooth boundary then it can be shown with the Divergence Theorem that  $\operatorname{per}_{\Omega}(\mathcal{I})$  coincides with the standard area of  $\partial\mathcal{I} \cap \Omega$ , i.e.,  $\mathcal{H}^{d-1}(\partial\mathcal{I} \cap \Omega)$ , the  $d-1$  dimensional Hausdorff measure (Rindler, 2018).

### 2.3. $\Gamma$ -convergence

$\Gamma$ -convergence is a tool for analyzing the limit behavior of a family of functionals. In our case we consider such a family, parameterized by  $\epsilon \downarrow 0$ :  $\mathcal{F}_{\epsilon} : L^1(\Omega) \rightarrow \mathbb{R}$ , and a “limit” functional  $\mathcal{F}_0 : L^1(\Omega) \rightarrow \mathbb{R}$ . We say that  $\mathcal{F}_{\epsilon}$   $\Gamma$ -converge to  $\mathcal{F}_0$  (as  $\epsilon \downarrow 0$ ) in the  $L^1(\Omega)$  norm, if the following two conditions hold:

1. *lim inf.* For all sequences  $u_{\epsilon} \rightarrow u$ ,

$$\mathcal{F}_0(u) \leq \liminf_{\epsilon \downarrow 0} \mathcal{F}_{\epsilon}(u_{\epsilon}).$$

2. *Recovery sequence.* For every  $v \in L^1(\Omega)$  there exists a sequence  $v_{\epsilon} \rightarrow v$  so that  $\lim_{\epsilon \downarrow 0} \mathcal{F}_{\epsilon}(v_{\epsilon}) = \mathcal{F}_0(v)$ .

The convergence of  $u_{\epsilon}, v_{\epsilon}$  in the above conditions is in the  $L^1(\Omega)$  sense. The most important property of  $\Gamma$ -convergence is that it implies convergence of minimizers and minimal values (Modica, 1987; Rindler, 2018):

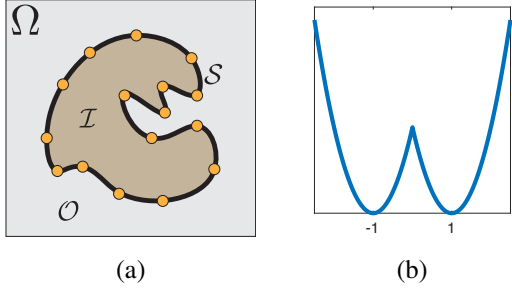


Figure 2. (a) The surface reconstruction problem asks to decompose a container ( $\Omega$ ) to interior ( $\mathcal{I}$ ) and exterior ( $\mathcal{O}$ ) with  $\mathcal{X}$  (centers of orange disks) being close to the interface  $\mathcal{S}$  (in black); (b) a visualization of a double-well potential  $W$ .

**Theorem 1.** *If  $u_\epsilon \rightarrow u_0$ , where  $u_\epsilon$  is a minimizer of  $\mathcal{F}_\epsilon$ , then  $\Gamma$ -convergence  $\mathcal{F}_\epsilon \rightarrow \mathcal{F}_0$  would imply  $u_0$  is a minimizer of  $\mathcal{F}_0$  and  $\lim_{\epsilon \downarrow 0} \mathcal{F}_\epsilon(u_\epsilon) = \mathcal{F}_0(u_0)$ .*

Furthermore, if the functional family  $\mathcal{F}$  is also *coercive*, meaning  $\mathcal{F}_\epsilon \leq c < \infty$ , for all  $\epsilon > 0$ , is compact, the convergence (up-to a subsequence) of  $u_\epsilon$  is established.

**Local minimizers.** While  $\Gamma$ -convergence deals with convergence of global minimizers, in case the limit functional  $\mathcal{F}_0$  has a strict local minimum  $u_0$ , the recovery sequence property shows that there exist a sequence of asymptotically local minimizers of  $\mathcal{F}_\epsilon$  that converge to  $u_0$ . In fact, there exists a recovery sequence of (perfect) local minimizers of  $\mathcal{F}_\epsilon$  to  $u_0$ , see Theorem 4.1 in (Kohn & Sternberg, 1989).

### 3. Method

We are considering the problem of learning an implicit neural representation (INR), i.e., equation 1, from an input raw geometry  $\mathcal{X} \subset \Omega$ , where  $\Omega$  is some Lipschitz domain (e.g., think of a bounding box to  $\mathcal{X}$ ). A prototypical example of raw geometry is a point cloud  $\mathcal{X} = \{\mathbf{x}_i\}_{i \in I}$ , where  $\mathbf{x}_i \in \mathbb{R}^d$ .

We are drawing inspiration from the phase transition problem of liquids (Modica, 1987; Sternberg, 1988; Gurtin, 1989; Hutchinson & Tonegawa, 2000; Rindler, 2018), where two liquids (e.g., oil and water) find equilibrium in a container ( $\Omega$ ) by minimizing a certain energy. In our case, the two liquids are chosen to represent the *interior* and the *exterior* of the shape. Mathematically, this is modeled with a function  $u : \Omega \rightarrow [-1, 1]$ . An  $x \in \Omega$  is deemed interior if  $u(x) = -1$ , exterior if  $u(x) = 1$ , and  $u(x) \in (-1, 1)$  represents some ‘‘mixture’’ of interior and exterior. We will denote by

$$\begin{aligned} \mathcal{I} &= \{\mathbf{x} \in \Omega \mid u(\mathbf{x}) \in [-1, 0)\}, \\ \mathcal{O} &= \{\mathbf{x} \in \Omega \mid u(\mathbf{x}) \in (0, 1]\}, \\ \mathcal{S} &= \partial\mathcal{I} \cap \partial\mathcal{O} \end{aligned} \quad (4)$$

the (rounded) interior of the shape, its exterior, and the boundary surface of the shape, respectively. In a sense,  $u$  is similar to occupancy functions used to supervise INRs (Mescheder et al., 2019; Chen & Zhang, 2019). In our case, however, we want to learn  $u$  directly from the raw data  $\mathcal{X}$ .

In phase transition of fluids one looks for an equilibrium density  $u$  minimizing the energy

$$\mathcal{E}(u) = \int_{\Omega} W(u), \quad (5)$$

over all  $u \in L^1(\Omega)$  satisfying the total mass constraint:

$$\int_{\Omega} u = m \in (-|\Omega|, |\Omega|). \quad (6)$$

The function  $W : \mathbb{R} \rightarrow \mathbb{R}$  is chosen to be a *double-well potential*, i.e., it is non-negative and has zeros only at  $\{-1, 1\}$ , e.g., Figure 2(b). As we will see, with this choice of  $W$ , the energy  $\mathcal{E}$  will drive (almost) every point  $\mathbf{x} \in \Omega$  to choose a side:  $u(\mathbf{x}) = -1$  (corresponding to the interior) or  $u(\mathbf{x}) = 1$  (corresponding to the exterior). The global mass constraint in equation 6 is motivated from the physical property that liquids maintain their mass in a closed container  $\Omega$ . This constraint, however, is not suitable for surface reconstruction since we don’t know the mass/volume of the shape we are looking for, and enforcing hard global constraints is challenging in deep networks, the main work horse of INRs.

We therefore start by considering the following (unconstrained) energy functional:

$$\mathcal{F}(u) = \lambda \mathcal{L}(u) + \begin{cases} \int_{\Omega} W(u) & u \in L^1(\Omega) \\ +\infty & \text{otherwise,} \end{cases} \quad (7)$$

where  $\lambda > 0$  is a parameter, and  $\mathcal{L}$  is the *reconstruction loss*, responsible for encouraging  $u$  to vanish in vicinity of  $\mathcal{X}$ :

$$\mathcal{L}(u) = \mathbb{E}_{\mathbf{x} \sim \mathcal{X}} \left| \int_{B(\mathbf{x})} u \right|, \quad (8)$$

where  $B(\mathbf{x})$  is some choice of a small radius ball centered at  $\mathbf{x}$ , and  $\mathbf{x} \sim \mathcal{X}$  means that  $\mathbf{x}$  is sampled from some distribution over  $\mathcal{X}$ . For example, if  $\mathcal{X}$  is a point cloud, a natural choice would be the uniform distribution over  $\mathcal{X}$ . We note that the following analysis also works for a more general reconstruction error: the integral in equation 8 can be replaced with  $\int_{\Omega} u \varphi_{\mathbf{x}}$ , where  $\varphi_{\mathbf{x}}(\mathbf{y})$  is, e.g., continuous and bounded in  $\mathbf{x}, \mathbf{y}$  over  $\Omega$ .

The energy  $\mathcal{L}$  can be seen as a local and soft version of the total density constrain in equation 6. It carries an interesting benefit that will be shown later on: in the context of the functional  $\mathcal{F}$  it will force  $u$  to change sign in vicinity of  $\mathcal{X}$ , rather than just zero out, which is a useful property for INRs.

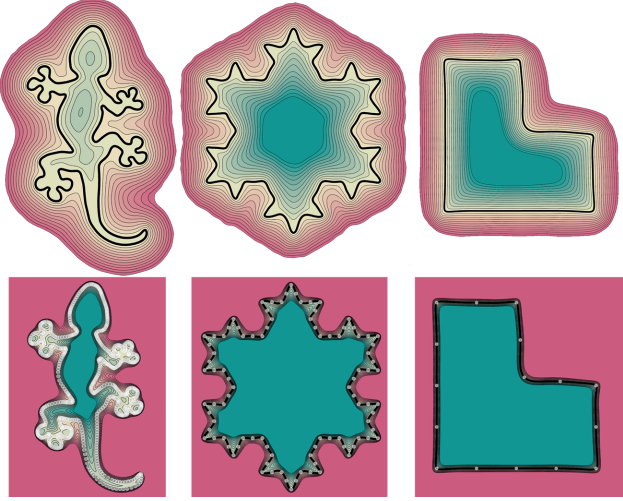


Figure 3. Learned densities  $u_\epsilon$  (bottom row) for input point cloud  $\mathcal{X}$  (depicted as white dots). The log transform of the densities,  $w_\epsilon$ , is visualized in the top row, with zero level set in bold. Negative values are colored in green, and positive in red.

It is not hard to see, however, that without further regularization equation 7 is too permissive, allowing far too many solutions (i.e., global minima):

**Lemma 2.** *Let  $\mathcal{X} = \{\mathbf{x}_i\}_{i \in I}$  be a (finite) point cloud, and assume the balls  $B$  are sufficiently small. Then, almost every arbitrary decomposition of the container  $\Omega = \mathcal{I} \cup \mathcal{O}$  so that  $\mathcal{X} \subset \partial\mathcal{I} \cap \Omega$  provides a global minimum to  $\mathcal{F}$ .*

*Proof.* Assume the balls  $B$  are sufficiently small so  $B(\mathbf{x}_i)$ ,  $i \in I$ , are all mutually disjoint. Next, consider an arbitrary decomposition  $\Omega = \mathcal{I} \cup \mathcal{O}$ , where the boundary  $\partial\mathcal{I}$  is smooth, and each  $B(\mathbf{x}_i) \cap \partial\mathcal{I}$  is exactly half a ball (i.e., congruent to  $B \cap$  half-space of  $\mathbb{R}^d$ ). Clearly, there is an infinite number of such decompositions. Lastly, let  $u(\mathbf{x}) = \mathbf{1}_{\mathcal{O}}(\mathbf{x}) - \mathbf{1}_{\mathcal{I}}(\mathbf{x})$ , where as above  $\mathbf{1}$  denotes the indicator function of a set. Then,  $\mathcal{F}(u) = 0$ .  $\square$

### 3.1. Regularization and minimal perimeter solutions

In phase transition theory the functional in equation 5 is further refined to enforce some regularity on the solutions  $u$  by incorporating  $\|\nabla u\|^2$ . This is called the Van der Waals-Cahn-Hilliard (WCH) theory of phase transitions (Modica, 1987; Sternberg, 1988). We adapt it for surface reconstruction:

$$\mathcal{F}_\epsilon(u) = \lambda \mathcal{L}(u) + \begin{cases} \int_\Omega \epsilon \|\nabla u\|^2 + W(u) & u \in W^{1,2}(\Omega) \\ +\infty & \text{otherwise} \end{cases} \quad (9)$$

and  $\epsilon > 0$  is a regularization parameter, and we switched to  $W^{1,2}(\Omega)$  Sobolev function space to allow integration of

the squared gradient norm; see Section 2.1 (and reference therein) for a short introduction on Sobolev spaces. Note that we added dependence on  $\epsilon$  in the functional notation; in our analysis we will consider  $\lambda$  to be a function of  $\epsilon$ , i.e.,  $\lambda = \lambda(\epsilon)$ .

As it turns out, the limit behavior of the minimizers of  $\mathcal{F}_\epsilon$ , denoted hereafter  $u_\epsilon$ , can be characterized pretty well, *regardless* of the specific choice of  $W$ : as  $\epsilon \downarrow 0$  they converge to a limit solution  $u_0$  with the following three key properties:

- (i) *Proper occupancy:*  $u_0 \in \text{BV}(\Omega; \{-1, 1\})$ .
- (ii) *Zero reconstruction error:*  $\mathcal{L}(u_0) = 0$ .
- (iii) *Minimal perimeter:*  $u_0$  minimizes  $\text{per}_\Omega(\mathcal{I})$ .

Several comments and clarifications are in order. First, (i) says that  $u_0$  takes values in  $\{-1, 1\}$  almost everywhere in  $\Omega$ , meaning it leaves no "mixed"/undetermined areas of positive measure; in other words, almost every point is either perfectly inside or outside the shape. Mathematically this means that  $u_0$  can be written as

$$u_0(\mathbf{x}) = -\mathbf{1}_{\mathcal{I}}(\mathbf{x}) + \mathbf{1}_{\Omega \setminus \mathcal{I}}(\mathbf{x}), \quad (10)$$

where as above  $\mathbf{1}$  is the indicator function of the relevant set,  $\mathcal{I}$  is defined in equation 4 with  $u = u_0$  (and in this case coincides up to measure zero with  $u_0^{-1}(-1)$ ), and the equality should be interpreted up to a set of measure zero. Second, property (ii),  $\mathcal{L}(u_0) = 0$ , means that  $u_0$  satisfies the constraints  $\int_{B(\mathbf{x})} u_0 = 0$  exactly, and in view of (i) indeed has to change sign in each  $B(\mathbf{x})$ . Third, since  $u_0 \in \text{BV}(\Omega; \{-1, 1\})$ , the set  $\mathcal{I}$  has a bounded surface perimeter in  $\Omega$  (see Section 2.2 and the supplementary). (iii) asserts that the  $\mathcal{I}$  will have minimal *surface perimeter*. Lastly, it can be shown that the set  $\mathcal{F}_\epsilon \leq c < \infty$  is compact in  $L^1(\Omega)$  (see e.g., Section 3.1 in (Kohn & Sternberg, 1989)), therefore convergence of  $u_\epsilon$  is up to a sub-sequence.

The three statements (i), (ii) and (iii) above are shown using the tool of  $\Gamma$ -convergence as introduced in Section 2. We define our target (limit) functional to be measuring the boundary surface perimeter for solutions that have zero reconstruction loss:

$$\mathcal{F}_0(u) = \begin{cases} \sigma_0 \text{per}_\Omega(\mathcal{I}) & u \in \text{BV}(\Omega, \{-1, 1\}), \\ & \text{and } \mathcal{L}(u) = 0 \\ +\infty & \text{otherwise} \end{cases} \quad (11)$$

where  $\sigma_0$  is a constant, and  $\mathcal{I}$  defined in equation 4. We prove:

**Theorem 2.** *If  $\lambda = \lambda(\epsilon)$  is chosen so that  $\lambda \epsilon^{-\frac{1}{2}} \rightarrow \infty$  while  $\lambda \epsilon^{-1/4} \rightarrow 0$ , then  $\epsilon^{-\frac{1}{2}} \mathcal{F}_\epsilon$   $\Gamma$ -converge to  $\mathcal{F}_0$ .*

The conditions on  $\lambda$  ask it to go to zero with  $\epsilon$  but at certain slower asymptotic rate; intuitively, letting  $\lambda = \epsilon^\alpha$ , then the conditions boil to  $\alpha \in (1/4, 1/2)$ . This is only a sufficient condition but gives a practical rough estimate for  $\lambda$  given  $\epsilon$  in equation 9.

Note that since  $\epsilon^{-\frac{1}{2}}\mathcal{F}_\epsilon$  and  $\mathcal{F}_\epsilon$  have the same minimizers, working in practice with  $\mathcal{F}_\epsilon$  (as we do) still enjoys convergence of minimizers. The proof is given in full in the supplementary. Here we provide the proof's idea.

*Proof idea.* The proof is based on  $\Gamma$ -convergence results of the WCH theory, e.g., (Modica, 1987; Sternberg, 1988; Fonseca & Tartar, 1988). We follow the proof technique in (Rindler, 2018), where the main novelty is the addition of the surface reconstruction loss  $\mathcal{L}$ . To prove  $\Gamma$ -convergence one must show that the two conditions formulated at Section 2.3, namely, the lim inf and recovery sequence, hold.

The lim inf is the relatively easier part of the WCH theory. We show that for an  $L^1(\Omega)$  convergent sequence  $u_\epsilon \rightarrow u$ ,  $\liminf_{\epsilon \downarrow 0} \mathcal{F}_\epsilon(u_\epsilon) \geq \mathcal{F}_0(u)$ . The main observation here is that the rate of  $\lambda = \lambda(\epsilon)$  forces the reconstruction loss to vanish in the limit, that is  $\mathcal{L}(u) = 0$ , showing property (ii) (property (i) comes out of the standard part of the WCH theory). After  $\mathcal{L}(u) = 0$  is shown the lim inf property comes out from existing arguments in the WCH theory.

The second part is establishing the recovery sequence, namely for every  $u \in L^1(\Omega)$  find a converging sequence in  $L^1(\Omega)$ ,  $u_\epsilon \rightarrow u$  where  $\mathcal{F}_\epsilon(u_\epsilon) \rightarrow \mathcal{F}_0(u)$ . The main observation here is that we can utilize the recovery sequence constructed in the WCH theory and show that it is also a recovery sequence for the reconstruction functionals. The idea is to show that if the reconstruction part satisfies  $\lambda(\epsilon)\mathcal{L}(u_\epsilon) \rightarrow 0$  then it does not interfere with the lim equality  $\mathcal{F}_\epsilon(u_\epsilon) \rightarrow \mathcal{F}_0(u)$ . This what leads to the asymptotic balance of  $\lambda$  required in the theorem formulation.

**Local minimizers.** Lastly, note that, as mentioned in Section 2.3,  $\Gamma$ -convergence would imply convergence of local minimizers to strict local minima of  $\mathcal{F}_0$ .

### 3.2. Signed distance functions

In this section we make use of the degree of freedom in choosing the double well potential  $W$  in equation 9, and show that a particular choice, together with a simple transformation of the minimizers  $u_\epsilon$  provide the fourth property:

(iv) *Signed distance function.*

In more detail, the signed distance function to the surface  $\mathcal{S}$  enclosing a shape  $\mathcal{I}$  is

$$d_{\mathcal{S}}(\mathbf{x}) = (-\mathbf{1}_{\mathcal{I}}(\mathbf{x}) + \mathbf{1}_{\Omega \setminus \mathcal{I}}(\mathbf{x})) \inf_{\mathbf{y} \in \mathcal{S}} \|\mathbf{x} - \mathbf{y}\|. \quad (12)$$

We advocate the following choice for the potential:

$$W(s) = s^2 - 2|s| + 1. \quad (13)$$

The graph of  $W$  is shown in Figure 2(b). Given a minimizer  $u_\epsilon$  of equation 9 we define our final INR by

$$w_\epsilon(\mathbf{x}) = -\sqrt{\epsilon} \log(1 - |u_\epsilon(\mathbf{x})|) \text{sign}(u_\epsilon(\mathbf{x})). \quad (14)$$

We claim that  $w_\epsilon$  will approximate  $d_{\mathcal{S}}$ . We first show:

**Theorem 3.** *Let  $u_\epsilon \in W^{1,2}(\Omega)$  be a (local) minimizer of  $\mathcal{F}_\epsilon$ , and  $O \subset \Omega \setminus \cup_{\mathbf{x} \in \mathcal{X}} B(\mathbf{x})$  a set where  $u_\epsilon \neq 0$ . Then,  $u_\epsilon$  is smooth in the classical sense in  $O$  and satisfies*

$$-\epsilon \Delta u_\epsilon + u_\epsilon - \text{sign}(u_\epsilon) = 0. \quad (15)$$

This theorem is proved by first applying the Euler-Lagrange conditions (see e.g., Theorem 3.1 in (Rindler, 2018)) for a minimizer  $u$  of  $\mathcal{F}_\epsilon$  in  $O$ , and noting that  $W'(s) = 2s - 2\text{sign}(s)$ , which is differentiable in  $O$ . In particular, equation 15 will be satisfied in the weak sense with any test function  $\psi \in C_c^\infty(O)$ . Second, regularity results for elliptic operators (e.g., Corollary 8.11 in (Gilbarg & Trudinger, 2015)) show that  $u$  is smooth in  $O$  and satisfies equation 15 in the classical sense. See the full details in the supplementary.

Theorem 3 can be used to show that  $w_\epsilon$  (as defined in equation 14) is a solution to the so-called *viscosity Eikonal equation*:

**Theorem 4.** *Let  $O \subset \Omega$  be a domain as defined in Theorem 3. Then, over  $O$ ,  $w_\epsilon$  satisfies*

$$-\sqrt{\epsilon} \Delta w_\epsilon + \text{sign}(u_\epsilon)(\|\nabla w_\epsilon\|^2 - 1) = 0 \quad (16)$$

The proof of this theorem is given in the supplementary material; the proof follows a direct computation. Note that the idea behind defining  $w_\epsilon$  is the so-called Cole-Hopf transformation, known for transforming some eikonal-type nonlinear equations to linear ones, see Section 4.4.1 in (Evans, 1998), and (Schieborn, 2006; Gurumoorthy & Rangarajan, 2009; Sethi et al., 2012; Belyaev & Fayolle, 2015).

Theorem 4 asserts  $w_\epsilon$  solves the viscosity Eikonal equation and it is expected, under certain conditions, to reproduce the signed distance function via the vanishing viscosity method (Crandall & Lions, 1983; Schieborn, 2006). We further show a closely related result by Varadhan (1967) allowing to prove that on some fixed domain  $O$ ,  $w_\epsilon$  defined by equations 14 and 15 converge to the signed distance function:

**Theorem 5.** *Let  $O$  be an open set, and  $u_\epsilon$  a solution to equation 15 in  $O$ ,  $u_\epsilon = 0$  on  $\partial O$  and  $u_\epsilon \neq 0$  in  $O$ . Then  $w_\epsilon \rightarrow \text{sign}(u_\epsilon)d_{\partial O}$  pointwise uniformly in any compact subset  $O \cup \partial O$ .*

Note that  $\text{sign}(u_\epsilon)$  is well defined in  $O$  since we assume that it does not vanish in  $O$ .

*Proof.* First, assume  $\text{sign}(u_\epsilon) > 0$  in  $O$ . Then,  $u_\epsilon$  satisfies:

$$\begin{aligned} -\epsilon \Delta u_\epsilon + u_\epsilon - 1 &= 0 && \text{in } O \\ u_\epsilon &= 0 && \text{in } \partial O \end{aligned}$$

Now the change of variables  $v_\epsilon = 1 - u_\epsilon$  leads to

$$\begin{aligned} \frac{1}{2} \Delta v_\epsilon &= \frac{1}{2\epsilon} v_\epsilon && \text{in } \Omega \\ v_\epsilon &= 1 && \text{in } \partial \Omega \end{aligned} \quad (17)$$

Therefore, Theorem 2.3 in (Varadhan, 1967) with  $\lambda = \frac{1}{2\epsilon}$  now implies

$$-\sqrt{\epsilon} \log(v_\epsilon) = -\sqrt{\epsilon} \log(1 - u_\epsilon) \xrightarrow{\epsilon \rightarrow 0} d_{\partial O}$$

uniformly in compact subsets of  $O \cup \partial O$ . The second case where  $\text{sign}(u_\epsilon) < 0$  in  $O$  is proved similarly, and provided in the supplementary.  $\square$

Using Theorem 5 to provide a rigorous proof of convergence of  $w_\epsilon$  requires two extra things: first, deal with the reconstructions constraints, equation 8, that inject another perturbation in vicinity of data points (but can be made arbitrary small), and establishing a version of Theorem 5 with slightly moving boundaries  $O_\epsilon$ . We leave these further analyses to future work.

### 3.3. Implementation details

**Additional losses.** Although not included in our analysis (and marked as interesting future work), the following reconstruction losses that include gradients of the learned INR are very useful in practice: If given as input a normal field over  $\mathcal{X}$ , namely  $\mathbf{n} : \mathcal{X} \rightarrow \mathcal{S}(\mathbb{R}^d)$ , where  $\mathcal{S}(\mathbb{R}^d)$  is the unit sphere, then the loss

$$\mathcal{N}_{\text{intr}}(u) = \mathbb{E}_{\mathbf{x} \sim \mathcal{X}} \left\| \mathbf{n}(\mathbf{x}) - \int_{B(\mathbf{x})} \nabla w \right\|^p \quad (18)$$

encourages the normal to  $\mathcal{S}$  to be close to  $\mathbf{n}$  over  $\mathcal{X}$ . Note that for  $\mathbf{x} \in \mathcal{S}$ , equation 14 implies that  $\nabla_{\mathbf{x}} w(\mathbf{x}) = \sqrt{\epsilon} \nabla_{\mathbf{x}} u(\mathbf{x})$ . In our implementation we used  $p = 1$ .

In cases the normal is not known but we want to incorporate more smoothness for  $\mathcal{S}$  at the data points, we can incorporate the unit gradient constraint, similar to (Gropp et al., 2020; Williams et al., 2020), only at data points  $\mathcal{S}$ :

$$\mathcal{N}_{\text{unit}}(u) = \mathbb{E}_{\mathbf{x} \sim \mathcal{X}} \left| 1 - \int_{B(\mathbf{x})} \|\nabla w\| \right|^p, \quad (19)$$

where we used  $p = 2$ , as in (Gropp et al., 2020).

**Network architecture and initialization.** We have used the same architecture for  $u$  as previous work (Park et al., 2019; Gropp et al., 2020), namely  $u : \mathbb{R}^d \rightarrow \mathbb{R}$  is a multi-layer perceptron (MLP) with 8 layer of 512 hidden neuron each, and a single skip connection concatenating the input vector to the input to the 4-th layer. We used either ReLU or Softplus activation with  $\beta = 100$ , and the geometric initialization of (Atzmon & Lipman, 2020a) initializing the network to be an approximated SDF to the  $d$ -dimensional unit sphere. We normalized  $\mathcal{X}$  to have unit max norm, and took  $\Omega$  to be a scaled version of the axis-aligned bounding box of  $\mathcal{X}$ ; we use as scale 1.5 in  $d = 3$ , and 2 in  $d = 2$ .

The loss we optimize for learning  $u$  is

$$\text{loss}(\theta) = \mathcal{F}_\epsilon + \mu \mathcal{N}_*, \quad (20)$$

where  $\theta$  are the parameters of the MLP  $u$ , and  $\mathcal{N}_*$  is defined either by equation 18 or equation 19, and  $\mu \geq 0$  is a hyperparameter. In total we have three hyper-parameters  $\epsilon, \lambda, \mu$ . In all the experiments we used  $\epsilon = 0.01$ , and did a parameter search over  $\lambda, \mu \in [0.2, 20]$  (note that  $\epsilon^{1/3} \approx 0.2$ , and  $1/3$  is in the range asserted by Theorem 2), including also  $\mu = 0$ . The integrals in the loss are stochastically estimated by sampling  $\mathbf{x} \sim \mathcal{U}(\Omega)$ , the uniform distribution over  $\Omega$ , and  $\mathbf{x} \sim \mathcal{N}(\mathbf{x}, \sigma^2)$ , the normal distribution centered at  $\mathbf{x}$  with standard deviation  $\sigma \in \{10^{-3}, 10^{-4}\}$ . In some experiments we used Fourier features (Tancik et al., 2020) as a first fixed layer in the network,  $\delta_k(\mathbf{x}) \in \mathbb{R}^{2kd}$ , where  $k$  is representing the number of frequencies used. See the supplementary for more implementation details.

## 4. Previous work

Representing shapes as zero level sets of neural networks has been popularized recently (Park et al., 2019; Mescheder et al., 2019; Chen & Zhang, 2019; Atzmon et al., 2019), and since been used for many applications in 3D vision and graphics. One standing challenge, tackled in this paper is learning INRs from raw geometric data. Similar problem was addressed in (Atzmon & Lipman, 2020a; Gropp et al., 2020; Sitzmann et al., 2020; Tancik et al., 2020; Atzmon & Lipman, 2020a; Williams et al., 2020). In IGR (Gropp et al., 2020) the authors use the implicit regularization property of neural network training to fit a smooth SDF to an input point cloud with or without normals. Williams et al. (2020) use the limit kernel of infinitely wide shallow networks to directly solve a kernel regression problem to fit an implicit representation to a point cloud with normals. High frequency signal learning techniques show that adding an input Fourier feature layer (Tancik et al., 2020), or using periodic activation functions (Sitzmann et al., 2020) facilitate faster learning of high frequency details. In this group of works the minimizers of the loss are generally not well understood and indeed, as we show in the experiments, sometimes lead to undesirable reconstructed parts. The minimal perimeter property seems to add an important inductive bias.

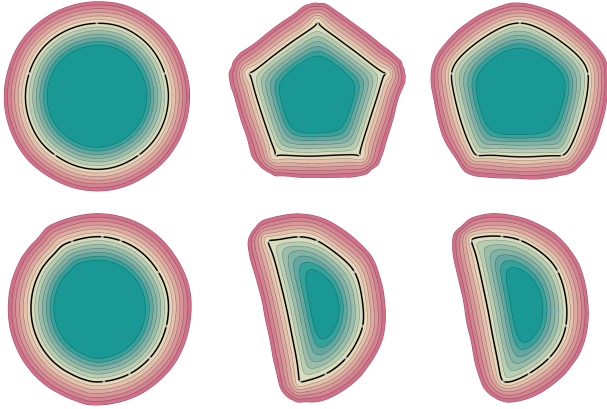


Figure 4. Reconstruction of 5-gon point cloud (top) and samples from half a circle (bottom). The left column shows the results of IGR, the middle columns shows PHASE without gradient loss, and right column is PHASE with gradient loss. Note the minimal perimeter property of the PHASE solutions.

In the signed agnostic learning (SAL) method (Atzmon & Lipman, 2020a;b) unsigned distance is fitted by a sign-agnostic regression to introduce a signed local minimum. In fact, SAL can be seen as a version of the unregularized loss in equation 7. Indeed, the loss  $||a| - b|^\ell$  (see equation 5 in (Atzmon & Lipman, 2020a)) is a double well potential as a function of  $a$ , similar to  $W$  in our formulation, with the difference of spatially changing  $W$ . In this sense our work suggests that adding a gradient regularization to SAL could be a useful idea. Generalizing our analysis to spatially varying potential, could also be interesting future research.

## 5. Experiments

We present experiments for  $d \in \{2, 3\}$  with the phase transition loss (PHASE) and compare to relevant baselines: IGR (Gropp et al., 2020), DGP (Williams et al., 2019), SIREN (Sitzmann et al., 2020), FFN (Tancik et al., 2020), NSP (Williams et al., 2020). For evaluation we use the same metrics as in (Williams et al., 2020): one-sided and double-sided Chamfer ( $d_C^{\rightarrow}(\mathcal{Y}_1, \mathcal{Y}_2)$ ,  $d_C(\mathcal{Y}_1, \mathcal{Y}_2)$ ) and Hausdorff distances ( $d_H^{\rightarrow}(\mathcal{Y}_1, \mathcal{Y}_2)$ ,  $d_H(\mathcal{Y}_1, \mathcal{Y}_2)$ ), see the supplementary for exact definitions.

### 5.1. 2D evaluation

Figure 3 depicts the PHASE learned densities  $u_\epsilon$  (bottom row) and their log transforms  $w_\epsilon$  defined in equation 14 (top row). The input point cloud in each example is depicted with white dots in the bottom row. In this example we use the loss in equation 20 with  $\mu = 0.1$ , and  $\lambda = 0.3$ . Note that indeed  $u_\epsilon$  approximates an occupancy function, while  $w_\epsilon$  approximates a signed distance function.

Figure 4 compares the reconstructions of IGR (left), and PHASE with  $\mu = 0$  (middle), and PHASE with  $\mu = 0.1$

Model	Method	Ground Truth		Scans	
		$d_C$	$d_H$	$d_C^{\rightarrow}$	$d_H^{\rightarrow}$
Anchor	DGP	0.33	8.82	0.08	2.79
	IGR	0.22	4.71	0.12	1.32
	SIREN	0.27	6.18	0.13	1.88
	FFN	0.31	4.49	0.10	0.10
	NSP	0.22	4.65	0.11	1.11
	PHASE	<b>0.21</b>	<b>4.29</b>	0.09	1.23
Daratech	DGP	0.20	3.14	0.04	1.89
	IGR	0.25	4.01	0.08	1.59
	SIREN	0.29	4.46	0.12	1.65
	FFN	0.34	5.97	0.10	0.10
	NSP	0.21	4.35	0.08	1.14
	PHASE	<b>0.18</b>	<b>2.92</b>	0.08	1.80
DC	DGP	0.18	4.31	0.04	2.53
	IGR	0.17	2.22	0.09	2.61
	SIREN	0.18	2.27	0.09	1.92
	FFN	0.20	2.87	0.10	0.12
	NSP	<b>0.14</b>	<b>1.35</b>	0.06	2.75
	PHASE	0.15	2.52	0.05	2.78
Gargoyle	DGP	0.21	5.98	0.06	3.41
	IGR	<b>0.16</b>	3.52	0.06	0.81
	SIREN	0.29	3.90	0.13	1.93
	FFN	0.22	5.04	0.09	0.09
	NSP	<b>0.16</b>	3.20	0.08	2.75
	PHASE	<b>0.16</b>	<b>3.14</b>	0.07	1.09
Lord Quas	DGP	0.14	3.67	0.04	2.03
	IGR	0.12	1.17	0.07	0.98
	SIREN	0.13	0.89	0.06	0.96
	FFN	0.35	3.90	0.06	0.06
	NSP	0.12	<b>0.69</b>	0.05	0.62
	PHASE	<b>0.11</b>	0.96	0.04	0.96

Table 1. Surface reconstruction results on the benchmark of (Williams et al., 2019).

(right), the input point cloud in each example is depicted with white dots; the zero level-set is in bold; in both cases  $\lambda = 0.3$ . Note that the PHASE results indeed exhibit smaller perimeter surface, while the IGR results tend to produce higher area extrapolation. Further note that incorporating the gradient loss  $\mathcal{N}_{\text{unit}}$  (with no normal data) increase the regularity of the reconstructed  $\mathcal{S}$  at the data points.

### 5.2. Surface reconstruction benchmark

We evaluated our loss on the dataset in (Williams et al., 2019), and compared to relevant baselines. This dataset consists of noisy range scan, including point cloud  $\mathcal{X}$  and normals  $\mathbf{n}$ , with complex geometry and topologies. Since normals are available we use in this case the loss in equation 20 with  $\mathcal{N}_{\text{intr}}$  from equation 18. We note that DGP already established itself as superior to many classical surface reconstruction methods. Furthermore, the results of the baselines are taken from NSP that performed parameter sweep for all methods (except itself, which is not an INR method per se) and chose individual best parameter for each model; we used the *same* set of parameters for all the models, i.e.,  $\lambda = \mu = 10$ . Results are summarized in Table 1.

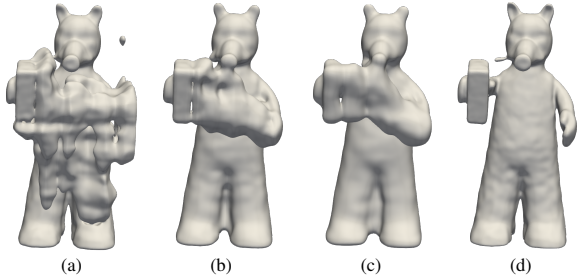


Figure 5. Reconstruction with no normals and Fourier features. (a-c) is IGR+FF, note the extraneous parts; (d) is PHASE+FF, removal of extraneous parts is attributed to the minimal perimeter.

Note that we achieve best reconstruction error in 4 out of 5 cases for Chamfer distance, and 3 out of 5 in Hausdorff distance.

### 5.3. Learning from points clouds and Fourier features

In this section we work with the surface reconstruction benchmark (see Section 5.2) but this time in the more challenging scenario of using only point cloud data (without normals). We compared PHASE with IGR together with high frequency learning that allow fast learning of details; we used Fourier features as described in Section 3.3 with  $k = 6$ . Similarly to previous work (Williams et al., 2020), we notice that in the point cloud only case the high frequency method possess an inductive bias that tends to introduce extraneous surface parts. For example, Figure 5 (a), (b), and (c) show the results of IGR training with Fourier features (FF); (d) shows the result of training PHASE with the same Fourier feature map. We attribute the reduction in the extraneous parts to the minimal surface perimeter property of our loss. Since this is a point only case, we trained equation 20 with equation 19; we note that in this case high  $\mu$  values cause some instability, and we conducted the experiment with  $\lambda = 10$ , and  $\mu = 0.5$ ; we present more results including also failure cases in the supplementary. Table 2 provides the quantitative success of PHASE and IGR with Fourier features on this benchmark; for IGR we did a parameter sweep and chose the parameters that achieved optimal  $d_C, d_H$  on the Gargoyle model. Note that PHASE is on par with some of the methods that use normal data in Section 5.2.

### 5.4. Large models

In this experiment we trained on point clouds and normals extracted from large models taken from the Stanford 3D Scanning Repository (Source: Stanford University Computer Graphics Laboratory). The point clouds consist of 0.5m-14m vertices, and were trained for 100k iteration of 1k batches with Fourier features  $k = 6$ . Figure 6 shows qualitative result of these reconstructions. Notably, the largest model, "Lucy" (left, 14m points) was trained for only 7 epochs and still presents relatively high level of details and

Model	Method	Ground Truth		Scans	
		$d_C$	$d_H$	$d_C^{\rightarrow}$	$d_H^{\rightarrow}$
Anchor	IGR+FF	0.72	9.48	0.24	8.89
	PHASE+FF	0.29	7.43	0.09	1.49
Daratech	IGR+FF	2.48	19.6	0.74	4.23
	PHASE+FF	0.35	7.24	0.08	1.21
DC	IGR+FF	0.86	10.32	0.28	3.98
	PHASE+FF	0.19	4.65	0.05	2.78
Gargoyle	IGR+FF	0.26	5.24	0.18	2.93
	PHASE+FF	0.17	4.79	0.07	1.58
Lord Quas	IGR+FF	0.49	10.71	0.14	3.71
	PHASE+FF	0.11	0.71	0.05	0.74

Table 2. Using Fourier features for fast training on the benchmark of (Williams et al., 2019) using only point data (no normals).



Figure 6. Reconstruction of large models from the Stanford 3D Scanning Repository.

no visible artifacts. We used the same parameters as in Section 5.2, i.e.,  $\lambda = \mu = 10$ .

## 6. Conclusions

We presented a new loss for surface reconstruction from raw data, suitable for the framework of implicit neural representations. This new loss learns a signed density function and is motivated from a mathematical physics model of liquids equilibrium in a container. This theory provides a principled tool to analyze the limit behavior of the loss minimizers. Most notably, convergence to occupancy function with minimal surface perimeter. By choosing a specific energy potential  $W$  we show that the log transform of the learned density approximates the signed distance function. Experiments with this loss demonstrate its usefulness for learning INRs from raw data. Interesting future work is to expand the analysis of the PHASE loss to include other reconstruction terms (e.g., equations 18 and 19); another interesting venue is to connect this loss to prior losses, possibly better understanding their limit behaviour and bias. A more ambitious goal would be to develop a general mathematical theory for INR losses, with well understood bias that can be tuned and even learned from data.



## References

- Adams, R. A. and Fournier, J. J. *Sobolev spaces*. Elsevier, 2003.
- Atzmon, M. and Lipman, Y. Sal: Sign agnostic learning of shapes from raw data. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020a.
- Atzmon, M. and Lipman, Y. Sal++: Sign agnostic learning with derivatives. *arXiv preprint arXiv:2006.05400*, 2020b.
- Atzmon, M., Haim, N., Yariv, L., Israelov, O., Maron, H., and Lipman, Y. Controlling neural level sets. In *Advances in Neural Information Processing Systems*, pp. 2032–2041, 2019.
- Belyaev, A. G. and Fayolle, P.-A. On variational and pde-based distance function approximations. In *Computer Graphics Forum*, volume 34, pp. 104–118. Wiley Online Library, 2015.
- Chen, Z. and Zhang, H. Learning implicit fields for generative shape modeling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5939–5948, 2019.
- Crandall, M. G. and Lions, P.-L. Viscosity solutions of hamilton-jacobi equations. *Transactions of the American mathematical society*, 277(1):1–42, 1983.
- Evans, L. *Partial differential equations (graduate studies in mathematics, v. 19, american mathematical society, providence, ri)*. 1998.
- Fonseca, I. and Tartar, L. The gradient theory of phase transitions for systems with two potential wells. 1988.
- Gilbarg, D. and Trudinger, N. S. *Elliptic partial differential equations of second order*. springer, 2015.
- Gropp, A., Yariv, L., Haim, N., Atzmon, M., and Lipman, Y. Implicit geometric regularization for learning shapes. In *Proceedings of Machine Learning and Systems 2020*, 2020.
- Gurtin, M. E. On phase transitions with bulk, interfacial, and boundary energy. In *Analysis and Continuum Mechanics*, pp. 429–450. Springer, 1989.
- Gurumoorthy, K. S. and Rangarajan, A. A schrödinger equation for the fast computation of approximate euclidean distance functions. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pp. 100–111. Springer, 2009.
- Hutchinson, J. E. and Tonegawa, Y. Convergence of phase interfaces in the van der waals-cahn-hilliard theory. *Calculus of Variations and Partial Differential Equations*, 10(1):49–84, 2000.
- Jiang, C., Sud, A., Makadia, A., Huang, J., Nießner, M., Funkhouser, T., et al. Local implicit grid representations for 3d scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6001–6010, 2020.
- Kohn, R. V. and Sternberg, P. Local minimisers and singular perturbations. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics*, 111(1-2):69–84, 1989.
- Leoni, G. *A first course in Sobolev spaces*. American Mathematical Soc., 2017.
- Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., and Geiger, A. Occupancy networks: Learning 3d reconstruction in function space. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4460–4470, 2019.
- Modica, L. The gradient theory of phase transitions and the minimal interface criterion. *Archive for Rational Mechanics and Analysis*, 98(2):123–142, 1987.
- Niemeyer, M., Mescheder, L., Oechsle, M., and Geiger, A. Occupancy flow: 4d reconstruction by learning particle dynamics. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5379–5389, 2019.
- Niemeyer, M., Mescheder, L., Oechsle, M., and Geiger, A. Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3504–3515, 2020.
- Park, J. J., Florence, P., Straub, J., Newcombe, R., and Lovegrove, S. DeepSDF: Learning continuous signed distance functions for shape representation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- Rindler, F. *Calculus of Variations*. Springer, 2018.
- Saito, S., Huang, Z., Natsume, R., Morishima, S., Kanazawa, A., and Li, H. Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2304–2314, 2019.
- Schieborn, D. Viscosity solutions of hamilton-jacobi equations of eikonal type on ramified spaces. 2006.

- Sethi, M., Rangarajan, A., and Gurumoorthy, K. The schrödinger distance transform (sdt) for point-sets and curves. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 198–205. IEEE, 2012.
- Sitzmann, V., Martel, J., Bergman, A., Lindell, D., and Wetzstein, G. Implicit neural representations with periodic activation functions. *Advances in Neural Information Processing Systems*, 33, 2020.
- Sternberg, P. The effect of a singular perturbation on nonconvex variational problems. *Archive for Rational Mechanics and Analysis*, 101(3):209–260, 1988.
- Tancik, M., Srinivasan, P. P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J. T., and Ng, R. Fourier features let networks learn high frequency functions in low dimensional domains. *NeurIPS*, 2020.
- Varadhan, S. R. S. On the behavior of the fundamental solution of the heat equation with variable coefficients. *Communications on Pure and Applied Mathematics*, 20(2):431–455, 1967.
- Williams, F., Schneider, T., Silva, C., Zorin, D., Bruna, J., and Panozzo, D. Deep geometric prior for surface reconstruction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 10130–10139, 2019.
- Williams, F., Trager, M., Bruna, J., and Zorin, D. Neural splines: Fitting 3d surfaces with infinitely-wide neural networks. *arXiv preprint arXiv:2006.13782*, 2020.
- Yariv, L., Kasten, Y., Moran, D., Galun, M., Atzmon, M., Ronen, B., and Lipman, Y. Multiview neural surface reconstruction by disentangling geometry and appearance. *Advances in Neural Information Processing Systems*, 33, 2020.

---

# Phase Transitions, Distance Functions, and Implicit Neural Representations

## Supplementary Material

---

### A. $\Gamma$ -convergence

*Proof (Theorem 2).* We will adapt the proof of Theorem 13.6 in (Rindler, 2018) to the surface reconstruction loss. We want to prove  $\Gamma$ -convergence of  $\epsilon^{-1/2}\mathcal{F}_\epsilon$  to  $\mathcal{F}_0$ . Note that  $\epsilon^{-1/2}\mathcal{F}_\epsilon$  is equivalent to the functional

$$\mathcal{F}_\epsilon(u) = \lambda\mathcal{L}(u) + \begin{cases} \int_\Omega \epsilon \|\nabla u\|^2 + \frac{1}{\epsilon}W(u) & u \in W^{1,2}(\Omega) \\ +\infty & \text{otherwise} \end{cases} \quad (21)$$

where now  $\lambda(\epsilon) = \epsilon^{-1}\tilde{\lambda}(\epsilon^2)$ , where  $\tilde{\lambda}$  represents the dependence in equation 9 (in the main paper). So we will prove  $\Gamma$ -convergence of this  $\mathcal{F}_\epsilon$  to  $\mathcal{F}_0$ , where  $\lambda \rightarrow \infty$  and  $\lambda\sqrt{\epsilon} \rightarrow 0$ , as  $\epsilon \downarrow 0$ . The proof of  $\Gamma$ -convergence requires showing the  $\liminf$  and *recovery* properties from Section 2.3 in the main paper. Let us denote:

$$\mathcal{E}_\epsilon(u) = \begin{cases} \int_\Omega \epsilon \|\nabla u\|^2 + \frac{1}{\epsilon}W(u) & u \in W^{1,2}(\Omega) \\ +\infty & \text{otherwise} \end{cases}$$

$$\mathcal{E}_0(u) = \begin{cases} \sigma_0 \text{per}_\Omega(\mathcal{I}) & u \in \text{BV}(\Omega; \{-1, 1\}) \\ +\infty & \text{otherwise} \end{cases}$$

For the reader's convenience we also repeat the definition of the limit functions  $\mathcal{F}_0$ :

$$\mathcal{F}_0(u) = \begin{cases} \sigma_0 \text{per}_\Omega(\mathcal{I}) & u \in \text{BV}(\Omega, \{-1, 1\}), \\ & \text{and } \mathcal{L}(u) = 0 \\ +\infty & \text{otherwise} \end{cases} \quad (22)$$

**Lim inf part.** In this part we need to consider  $u_\epsilon \rightarrow u$  in  $L^1(\Omega)$ ; we abuse notation a bit and let  $\epsilon \downarrow 0$  denote some particular sequence  $\epsilon_k \rightarrow 0$  as  $k \rightarrow \infty$ . We need to show that  $\liminf_{\epsilon \downarrow 0} \mathcal{F}_\epsilon(u_\epsilon) \geq \mathcal{F}_0(u)$ . If  $\liminf_{\epsilon \downarrow 0} \mathcal{F}_\epsilon(u_\epsilon) = \infty$  the statement holds trivially, therefore we assume  $\liminf_{\epsilon \downarrow 0} \mathcal{F}_\epsilon(u_\epsilon) < \infty$ .

In the proof of Theorem 13.6 in Rindler (2018) it is shown that  $\int_\Omega W(u(\mathbf{x}))d\mathbf{x} = 0$  and consequently  $u(\mathbf{x}) \in \{-1, +1\}$  almost everywhere, i.e., equation 10 (in the main paper) holds for  $u$ . Furthermore, it is shown that

$$\liminf_{\epsilon \downarrow 0} \mathcal{E}_\epsilon(u_\epsilon) \geq \mathcal{E}_0(u), \quad (23)$$

where  $\sigma_0$  is some constant depending on  $W$  alone:

$$\sigma_0 = 2 \int_{-1}^1 \sqrt{W(s)}ds. \quad (24)$$

Since  $\mathcal{E}_0(u) < \infty$  in particular means that  $u \in \text{BV}(\Omega; \{-1, 1\})$  (this can be seen directly from the definitions in equations 2 and 3 in the main paper).

Now, for our reconstruction loss, since  $u_\epsilon \rightarrow u$  in  $L^1(\Omega)$  we have that for every  $\mathbf{x} \in \mathcal{X}$ ,  $\left| \int_{B(\mathbf{x})} u_\epsilon \right| \rightarrow \left| \int_{B(\mathbf{x})} u \right|$ , as  $\epsilon \downarrow 0$ .

By Fatou's lemma applied for the functions  $\mathbf{x} \mapsto \left| \int_{B(\mathbf{x})} u_\epsilon \right|$  and the limit function  $\mathbf{x} \mapsto \left| \int_{B(\mathbf{x})} u \right|$ , and the fact that  $\liminf_{\epsilon \downarrow 0} \mathcal{F}_\epsilon(u_\epsilon) < \infty$ :

$$\mathcal{L}(u) = \mathbb{E}_{\mathbf{x}} \left| \int_{B(\mathbf{x})} u \right| \leq \liminf_{\epsilon \downarrow 0} \mathbb{E}_{\mathbf{x}} \left| \int_{B(\mathbf{x})} u_\epsilon \right|$$

$$= \liminf_{\epsilon \downarrow 0} \mathcal{L}(u_\epsilon) \leq \liminf_{\epsilon \downarrow 0} \frac{1}{\lambda} \mathcal{F}_\epsilon(u_\epsilon) = 0.$$

where the last equality is due to  $\lambda \rightarrow \infty$  as  $\epsilon \downarrow 0$ . This means that the limit function  $u$  satisfies the reconstruction constraints perfectly, or in other words that the reconstructed surface  $\mathcal{S}$  passes through all the balls  $B(\mathbf{x})$ ,  $\mathbf{x} \in \mathcal{X}$ , except possibly a subset of  $\mathcal{X}$  of measure zero. However, since  $\left| \int_{B(\mathbf{x})} u \right|$  is continuous as a function of  $\mathbf{x}$  this is true for all balls  $B(\mathbf{x})$ ,  $\mathbf{x} \in \mathcal{X}$ .

In particular  $\mathcal{E}_0(u) = \mathcal{F}_0(u)$ . Now incorporating this with equation 23 we get

$$\liminf_{\epsilon \downarrow 0} \mathcal{F}_\epsilon(u_\epsilon) = \liminf_{\epsilon \downarrow 0} (\lambda\mathcal{L}(u_\epsilon) + \mathcal{E}_\epsilon(u_\epsilon))$$

$$\geq \liminf_{\epsilon \downarrow 0} \mathcal{E}_\epsilon(u_\epsilon)$$

$$\geq \mathcal{E}_0(u) = \mathcal{F}_0(u)$$

as required.

**Recovery sequence part.** In this part we need to consider an arbitrary  $u \in L^1(\Omega)$  and find a sequence  $u_\epsilon \in L^1(\Omega)$  so that  $u_\epsilon \rightarrow u$  in  $L^1(\Omega)$  and  $\lim_{\epsilon \downarrow 0} \mathcal{F}_\epsilon(u_\epsilon) = \mathcal{F}_0(u)$ .

Let  $u \in L^1(\Omega)$  be arbitrary. If  $u \notin \text{BV}(\Omega; \{-1, 1\})$  or  $\mathcal{L}(u) > 0$  then  $\mathcal{F}_0(u) = \infty$  and there is no need to construct a recovery sequence in this case (see, e.g., Theorem 1 where no recovery sequence for such cases is needed). So we assume  $\mathcal{F}_0(u) < \infty$ , and  $u$  of the form  $u = -\mathbf{1}_{\mathcal{I}} + \mathbf{1}_{\Omega \setminus \mathcal{I}}$ , where  $\mathcal{I}$  is defined as in equation 4.

Next, note that if  $\int_\Omega u \in \{-|\Omega|, |\Omega|\}$  then  $\mathcal{L}(u) > 0$  and again  $\mathcal{F}_0(u) = \infty$ . Therefore we can assume  $\int_\Omega u \in (-|\Omega|, |\Omega|)$ . In this case, Theorem 13.6 in Rindler (2018)

shows the existence of a recovery sequence of functions  $u_\epsilon$  so that  $u_\epsilon \rightarrow u$  in  $L^1(\Omega)$  and  $\lim_{\epsilon \downarrow 0} \mathcal{E}_\epsilon(u_\epsilon) = \mathcal{E}_0(u)$ . The main observation in this part is that  $u_\epsilon$  is a recovery sequence also for our surface reconstruction functionals  $\mathcal{F}_\epsilon$  and  $\mathcal{F}_0$ .

To show that  $u_\epsilon$  is a recovery sequence also in our settings it is enough to show that  $\lambda\mathcal{L}(u_\epsilon) \rightarrow 0$  as  $\epsilon \downarrow 0$ . Indeed, if this is the case,

$$\begin{aligned} \lim_{\epsilon \downarrow 0} \mathcal{F}_\epsilon(u_\epsilon) &= \lim_{\epsilon \downarrow 0} (\mathcal{E}_\epsilon(u_\epsilon) + \lambda\mathcal{L}(u_\epsilon)) \\ &= \mathcal{E}_0(u) = \mathcal{F}_0(u) \end{aligned}$$

To show that  $\lambda\mathcal{L}(u_\epsilon) \rightarrow 0$  we need to use a bit of extra information on  $u_\epsilon$ :  $u_\epsilon$  is constructed to approximate  $w = -\mathbf{1}_{G \cap \Omega} + \mathbf{1}_{\Omega \setminus G}$ , that is

$$\int_{\Omega} |u_\epsilon - w| \rightarrow 0,$$

where  $G \subset \mathbb{R}^d$  is open, bounded with smooth boundary, and  $|(G \Delta \mathcal{I}) \cap \Omega|$  can be made arbitrary small (see Lemma 13.7 in Rindler (2018), or Lemma 1 in Modica (1987)). This means that for arbitrary  $\delta > 0$  we can choose  $G$  so that

$$\int_{\Omega} |u - w| \leq |(G \Delta \mathcal{I}) \cap \Omega| \leq \delta$$

Remember that  $\mathcal{L}(u) = 0$  and using this last inequality we get that for arbitrary  $\mathbf{x} \in \mathcal{X}$ ,

$$\begin{aligned} \left| \int_{B(\mathbf{x})} u_\epsilon \right| &= \left| \int_{B(\mathbf{x})} u_\epsilon - \int_{B(\mathbf{x})} w \right| + \left| \int_{B(\mathbf{x})} w - \int_{B(\mathbf{x})} u \right| \\ &\leq \int_{\Omega} |u_\epsilon - w| + \int_{\Omega} |w - u| \leq c\sqrt{\epsilon} \end{aligned} \quad (25)$$

where  $c > 0$  is some constant, and the last inequality is due to the fact that we can make the choice  $\delta = \sqrt{\epsilon}$  and the following bound shown in the proof of Theorem 13.6 in Rindler (2018):

$$\int_{\Omega} |u_\epsilon - w| \leq 4\sqrt{\epsilon} \sup_{-2\sqrt{\epsilon} \leq t \leq 2\sqrt{\epsilon}} \mathcal{H}^{d-1}(\mathcal{S}_t \cap \Omega),$$

where  $\mathcal{S}_t = \{\mathbf{x} \in \mathbb{R}^d \mid d_{\mathcal{S}}(\mathbf{x}) = t\}$ ,  $d_{\mathcal{S}}$  is the signed distance function defined in equation 12,  $\mathcal{S}$  is defined as in equation 4 for  $w$ , and Lemma 13.9 in Rindler (2018) shows that the Hausdorff measure of  $\mathcal{S}_t \cap \Omega$  converges to that of  $\partial\mathcal{I} \cap \Omega$  as  $\epsilon \downarrow 0$ , and therefore is bounded.

Equation 25 implies that

$$\lambda\mathcal{L}(u_\epsilon) \leq c\lambda\sqrt{\epsilon}.$$

Lastly, remember that  $\lambda\sqrt{\epsilon} \rightarrow 0$  and therefore  $u_\epsilon$  is a recovery sequence as desired.  $\square$

## B. Distance functions

**Theorem 3.** *Let  $u_\epsilon \in W^{1,2}(\Omega)$  be a (local) minimizer of  $\mathcal{F}_\epsilon$ , and  $O \subset \Omega \setminus \cup_{\mathbf{x} \in \mathcal{X}} B(\mathbf{x})$  a set where  $u_\epsilon \neq 0$ . Then,  $u_\epsilon$  is smooth in the classical sense in  $O$  and satisfies*

$$-\epsilon\Delta u_\epsilon + u_\epsilon - \text{sign}(u_\epsilon) = 0. \quad (26)$$

*Proof.* We start by applying the Euler-Lagrange (EL) conditions (see e.g., Theorem 3.1 in (Rindler, 2018)) for a minimizer  $u$  of  $\mathcal{F}_\epsilon$  in  $O$ . Let

$$f(x, v, \mathbf{v}) = \epsilon \|\mathbf{v}\|^2 + v^2 - 2|v| + 1,$$

be our integrand. That is  $\mathcal{F}_\epsilon(u) = \int_{\Omega} f(x, u, \nabla u)$ . The EL conditions are:

$$-\text{div} [\nabla_{\mathbf{v}} f(x, u, \nabla u)] + \frac{d}{dv} f(x, u, \nabla u) = 0.$$

Plugging our  $f$  and noting that  $W'(s) = 2s - 2\text{sign}(s)$ , which is differentiable in  $O$ , we get

$$-2\epsilon\Delta u + 2u - 2\text{sign}(u) = 0.$$

In particular, equation 26 will be satisfied in the weak sense with any test function  $\psi \in C_c^\infty(O)$ . Second, regularity results for elliptic operators (e.g., Corollary 8.11 in (Gilberg & Trudinger, 2015)) show that  $u$  is smooth in  $O$  and satisfies equation 26 in the classical sense.  $\square$

**Theorem 4.** *Let  $O \subset \Omega$  be a domain as defined in Theorem 3. Then, over  $O$ ,  $w_\epsilon$  satisfies*

$$-\sqrt{\epsilon}\Delta w_\epsilon + \text{sign}(u_\epsilon)(\|\nabla w_\epsilon\|^2 - 1) = 0 \quad (27)$$

*Proof.* For brevity, we denote  $u = u_\epsilon$ , and assume  $u > 0$  in  $O$ . Then, equation 14 (in the main paper) in this case is

$$w = -\sqrt{\epsilon} \log(1 - u).$$

Let  $\mathbf{x} = (x_1, x_2, \dots, x_d)$ . Now,

$$\begin{aligned} \frac{\partial w}{\partial x_i} &= \sqrt{\epsilon} \frac{1}{1-u} \frac{\partial u}{\partial x_i} \\ \frac{\partial^2 w}{\partial x_i^2} &= \sqrt{\epsilon} \frac{1}{(1-u)^2} \left[ \frac{\partial u}{\partial x_i} \right]^2 + \sqrt{\epsilon} \frac{1}{1-u} \frac{\partial^2 u}{\partial x_i^2} \end{aligned}$$

Plugging in the l.h.s. of equation 27 we get

$$\frac{-\epsilon}{(1-u)^2} \|\nabla u\|^2 + \frac{-\epsilon}{1-u} \Delta u + \frac{\epsilon}{(1-u)^2} \|\nabla u\|^2 - 1$$

and this term vanishes in view of equation 26.

If  $u < 0$  in  $O$ , then equation 14 (in the main paper) in this case is

$$w = \sqrt{\epsilon} \log(1 + u).$$

Similar to before:

$$\begin{aligned}\frac{\partial w}{\partial x_i} &= \sqrt{\epsilon} \frac{1}{1+u} \frac{\partial u}{\partial x_i} \\ \frac{\partial^2 w}{\partial x_i^2} &= -\sqrt{\epsilon} \frac{1}{(1+u)^2} \left[ \frac{\partial u}{\partial x_i} \right]^2 + \sqrt{\epsilon} \frac{1}{1+u} \frac{\partial^2 u}{\partial x_i^2}\end{aligned}$$

Again, plugging in the l.h.s. of equation 27 we get

$$\frac{\epsilon}{(1+u)^2} \|\nabla u\|^2 + \frac{-\epsilon}{1+u} \Delta u + 1 - \frac{\epsilon}{(1+u)^2} \|\nabla u\|^2$$

that again vanishes in view of equation 26.  $\square$

**Theorem 5.** *Let  $O$  be an open set, and  $u_\epsilon$  a solution to equation 26 in  $O$ ,  $u_\epsilon = 0$  on  $\partial O$  and  $u_\epsilon \neq 0$  in  $O$ . Then  $w_\epsilon \rightarrow \text{sign}(u_\epsilon) d_{\partial O}$  pointwise uniformly in any compact subset  $O \cup \partial O$ .*

Note that  $\text{sign}(u_\epsilon)$  is well defined in  $O$  since we assume that  $O$  does not vanish in  $O$ .

*Proof.* First, assume  $\text{sign}(u_\epsilon) > 0$  in  $O$ . Then,  $u_\epsilon$  satisfies:

$$\begin{aligned}-\epsilon \Delta u_\epsilon + u_\epsilon - 1 &= 0 && \text{in } O \\ u_\epsilon &= 0 && \text{in } \partial O\end{aligned}$$

Now the change of variables  $v_\epsilon = 1 - u_\epsilon$  leads to

$$\begin{aligned}\frac{1}{2} \Delta v_\epsilon &= \frac{1}{2\epsilon} v_\epsilon && \text{in } \Omega \\ v_\epsilon &= 1 && \text{in } \partial \Omega\end{aligned} \quad (28)$$

Therefore, Theorem 2.3 in (Varadhan, 1967) with  $\lambda = \frac{1}{2\epsilon}$  now implies

$$-\sqrt{\epsilon} \log(v_\epsilon) = -\sqrt{\epsilon} \log(1 - u_\epsilon) \xrightarrow{\epsilon \rightarrow 0} d_{\partial O}$$

uniformly in compact subsets of  $O \cup \partial O$ .

In the case  $\text{sign}(u_\epsilon) < 0$  in  $O$ ,  $u_\epsilon$  satisfies:

$$\begin{aligned}-\epsilon \Delta u_\epsilon + u_\epsilon + 1 &= 0 && \text{in } O \\ u_\epsilon &= 0 && \text{in } \partial O\end{aligned}$$

The change of coordinates  $v_\epsilon = u_\epsilon + 1$  now leads again to  $v_\epsilon$  satisfying equation 28 and invoking (Varadhan, 1967) again implies that

$$\sqrt{\epsilon} \log(v_\epsilon) = \sqrt{\epsilon} \log(1 + u_\epsilon) \xrightarrow{\epsilon \rightarrow 0} -d_{\partial O}$$

uniformly in compact subsets of  $O \cup \partial O$ . Putting the two cases together and comparing to equation 14 (in the main paper) proves the theorem.  $\square$

## C. Networks in Sobolev spaces

Let  $f : \mathbb{R}^d \times \mathbb{R}^p \rightarrow \mathbb{R}$  be a multilayer perceptron (MLP) with ReLU activation. That is,  $f(\mathbf{x}; \theta)$  is composed of layers of the form  $\mathbf{z} = \sigma(\mathbf{W}\mathbf{y} + \mathbf{b})$ , where  $\mathbf{W}$ ,  $\mathbf{b}$  are the parameters, collectively defining  $\theta \in \mathbb{R}^p$ , and  $\sigma(s) = \max\{0, s\}$  is the ReLU applied entry-wise. We consider the functions  $f(\cdot; \theta)$ , for a fixed  $\theta$ : these are piecewise linear and continuous. In fact, each linear piece  $L_k = f|_{\Omega_k}$  is defined over a polytope  $\Omega_k \subset \mathbb{R}^d$ . For the analysis of the loss functions we discuss in this paper we require a complete function space that contains these neural functions and allow discussing their "derivatives" and convergence. The most natural such space is  $W^{1,p}(\Omega)$ ,  $p \in [1, \infty)$ , the Sobolev space of all  $L^p(\Omega)$  functions with first weak derivatives also in  $L^p(\Omega)$ . We note that MLP with Softplus activation is smooth in the classical sense hence in particular belongs to  $W^{1,p}(\Omega)$ .

**Lemma 1.** *The functions  $f(\cdot; \theta)$ , where  $f$  is an MLP with ReLU activation belongs to  $W^{1,p}(\Omega)$ , for all  $p \in [1, \infty]$ , and  $\Omega$  Lipschitz domain.*

*Proof.* Denote  $f(\mathbf{x}) := f(\mathbf{x}; \theta)$ . To show that  $f \in W^{1,p}(\Omega)$  we need to show two things:  $f \in L^p(\Omega)$  (which is clear since  $f$  is measurable and bounded as well as  $\Omega$  is bounded), and that it has *weak derivatives* in  $L^p(\Omega)$ , that is for every  $i \in [d]$  there exists a function  $g \in L^p(\Omega)$  so that

$$\int_{\Omega} f \frac{\partial \psi}{\partial x_i} = - \int_{\Omega} g \psi, \quad (29)$$

for all compactly supported smooth functions  $\psi \in C_c^\infty(\Omega)$ . Let  $\Omega = \cup_k \Omega_k$  be the decomposition of  $\Omega$  to the subdomains where  $L_k = f|_{\Omega_k}$  is linear. The multivariate integration by parts formula provides:

$$\int_{\Omega_k} f \frac{\partial \psi}{\partial x_i} = \int_{\partial \Omega_k} f \psi \langle \mathbf{e}_i, \mathbf{n} \rangle - \int_{\Omega_k} \frac{\partial f}{\partial x_i} \psi. \quad (30)$$

Therefore, a natural candidate for the weak derivative  $\frac{\partial f}{\partial x_i}$  is the piecewise constant  $g$  that equals  $\frac{\partial L_k}{\partial x_i}$  in the interior of each  $\Omega_k$ . Indeed, noticing that  $\langle \mathbf{e}_i, \mathbf{n} \rangle$  flips sign when the normal  $\mathbf{n}$  flips sign, we see that all interior contributions of the boundary integrals  $\int_{\partial \Omega_k} f \psi \langle \mathbf{e}_i, \mathbf{n} \rangle$  cancel, and since  $\psi$  vanishes on the boundary  $\partial \Omega$ , summing equation 30 over  $k$  we get equation 29.  $\square$

## D. Experiments and implementation details

In all the experiments we used  $\epsilon = 0.01$ , and did a parameter search over  $\lambda \in \{0.2, 0.3, 0.5, 1, 10, 20\}$  (note that  $\epsilon^{1/3} \approx 0.2$ , and  $1/3$  is in the range suggested by Theorem 2), and similarly for  $\mu$ . The gradients  $\nabla u(\mathbf{x})$  are computed with automatic differentiation. A single training iteration with batch size  $16k$  takes 0.16sec, on an NVIDIA Quadro GP100.

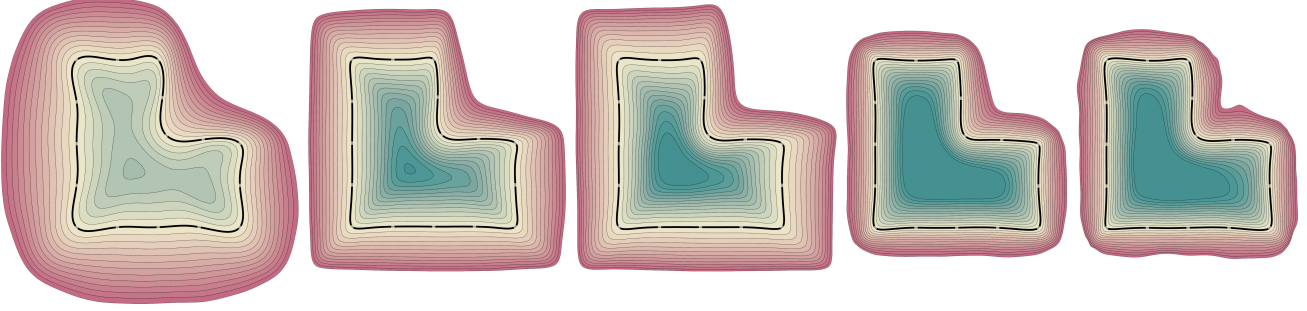


Figure 7. Ablation of  $\epsilon$ . From left to right  $\epsilon = 1.0, 0.1, 0.05, 0.01, 0.005$ ;  $\lambda = c\epsilon^{0.3}$  according to Theorem 2, and  $c \in \{1, 10\}$ .

### D.1. Fourier features

In some experiments we used Fourier features (Tancik et al., 2020) as the first constant layer in the network,  $\delta_k : \mathbb{R}^d \rightarrow \mathbb{R}^{2kd}$ , where  $k$  is a parameter representing the number of frequencies used. For  $\mathbf{x} \in \mathbb{R}^d$  the Fourier feature layer is defined to be a vector  $\delta_k(\mathbf{x}) \in \mathbb{R}^{2kd}$  with the real and imaginary parts of  $\exp(i2^\omega \pi x_j)$ ,  $\omega \in [k]$ ,  $j \in [d]$  as entries.

### D.2. Metrics

The distance between two point clouds  $\mathcal{X}_1, \mathcal{X}_2$  is computed as in (Williams et al., 2019) using the standard one-sided and double-sided Chamfer and Hausdorff distances. We denote by  $d_C^\rightarrow(\mathcal{X}_1, \mathcal{X}_2)$ ,  $d_C(\mathcal{X}_1, \mathcal{X}_2)$  the one-sided and double-sided  $\ell_1$  Chamfer distance; and by  $d_H^\rightarrow(\mathcal{X}_1, \mathcal{X}_2)$ ,  $d_H(\mathcal{X}_1, \mathcal{X}_2)$  the one-sided and double-sided Hausdorff distance. These are defined as follows:

$$\begin{aligned} d_C^\rightarrow(\mathcal{X}, \mathcal{Y}) &= \frac{1}{|\mathcal{X}|} \sum_{\mathbf{x} \in \mathcal{X}} \min_{\mathbf{y} \in \mathcal{Y}} \|\mathbf{x} - \mathbf{y}\|_2 \\ d_C(\mathcal{X}, \mathcal{Y}) &= \frac{1}{2} (d_C^\rightarrow(\mathcal{X}, \mathcal{Y}) + d_C^\rightarrow(\mathcal{Y}, \mathcal{X})) \\ d_H^\rightarrow(\mathcal{X}, \mathcal{Y}) &= \max_{\mathbf{x} \in \mathcal{X}} \min_{\mathbf{y} \in \mathcal{Y}} \|\mathbf{x} - \mathbf{y}\|_2 \\ d_H(\mathcal{X}, \mathcal{Y}) &= \max \{d_H^\rightarrow(\mathcal{X}, \mathcal{Y}) + d_H^\rightarrow(\mathcal{Y}, \mathcal{X})\}, \end{aligned}$$

To measure the distance between a surface and a point cloud, or between two surfaces, we first sample each surface  $\mathcal{S}$  densely (i.e., with  $10m$  uniformly random points)  $\mathcal{Y} \subset \mathcal{S}$ , and then measure the distance between the corresponding points clouds as described above.

### D.3. 2D evaluation

Figure 7 demonstrates ablation of  $\epsilon$ ; we show  $w_\epsilon$  for  $\epsilon$  in a range of values  $\epsilon \in \{1.0, 0.1, 0.05, 0.01, 0.005\}$ , where  $\lambda$  is chosen according to Theorem 2, i.e.,  $\lambda = c\epsilon^{0.3}$ , and we take  $c \in \{1, 10\}$ . Note that as  $\epsilon$  decreases  $w_\epsilon$  is closer to a distance function.

### D.4. Surface reconstruction benchmark

This benchmark consists of 5 models, the input (train data) to each model is a point cloud  $\mathcal{X}$  of size  $170k - 290k$ , and corresponding normal data  $\mathbf{n} : \mathcal{X} \rightarrow \mathcal{S}(\mathbb{R}^3)$ , as well as ground-truth test point cloud  $\mathcal{Y}$ . In this experiment we used the PHASE loss in equation 20 (in the main paper) with the normal loss equation 18 (in the main paper), and  $\mu = \lambda = 10$ . The batch size was taken to be  $16k$  and we performed  $100k$  iterations, which roughly correspond to  $5k - 10k$  epochs. Figure 8 shows qualitative results of PHASE.

**PHASE and Fourier Features.** Training PHASE with Fourier features and normal data leads to comparable results to PHASE alone, but requires order of magnitude less iterations. In Table 3 we show the results of PHASE trained with Fourier Features ( $k = 6$ ), on the surface reconstruction benchmark of (Williams et al., 2019) (i.e., point clouds with normals), with the same parameters  $\mu = \lambda = 10$ , as before. In this case we trained the model for  $10k$  iterations with batch-size  $16k$ . See Table 1 in the main paper for comparison with other methods. Note that PHASE and PHASE+FF are roughly equivalent according to the distance metric scores, maybe with slight advantage to PHASE without FF. However, PHASE+FF was trained for  $10k$  iterations versus  $100k$  iterations in the case of PHASE without FF, and it also exhibits slightly more high frequency details not captured by the metric scores, see Figure 8.

### D.5. Learning from point clouds and Fourier features

In this example we again worked with the surface reconstruction benchmark but explored the more challenging case of removing the normal data and using only the point clouds  $\mathcal{X}$ . We wanted to see the bias caused by the high frequency methods and compared PHASE and IGR in the case of adding Fourier features  $k = 6$  to both. For PHASE we used  $\lambda = 10$  as before, but smaller  $\mu$ , i.e.,  $\mu = 0.5$ , as we found larger  $\mu$  is more unstable. For IGR we did a

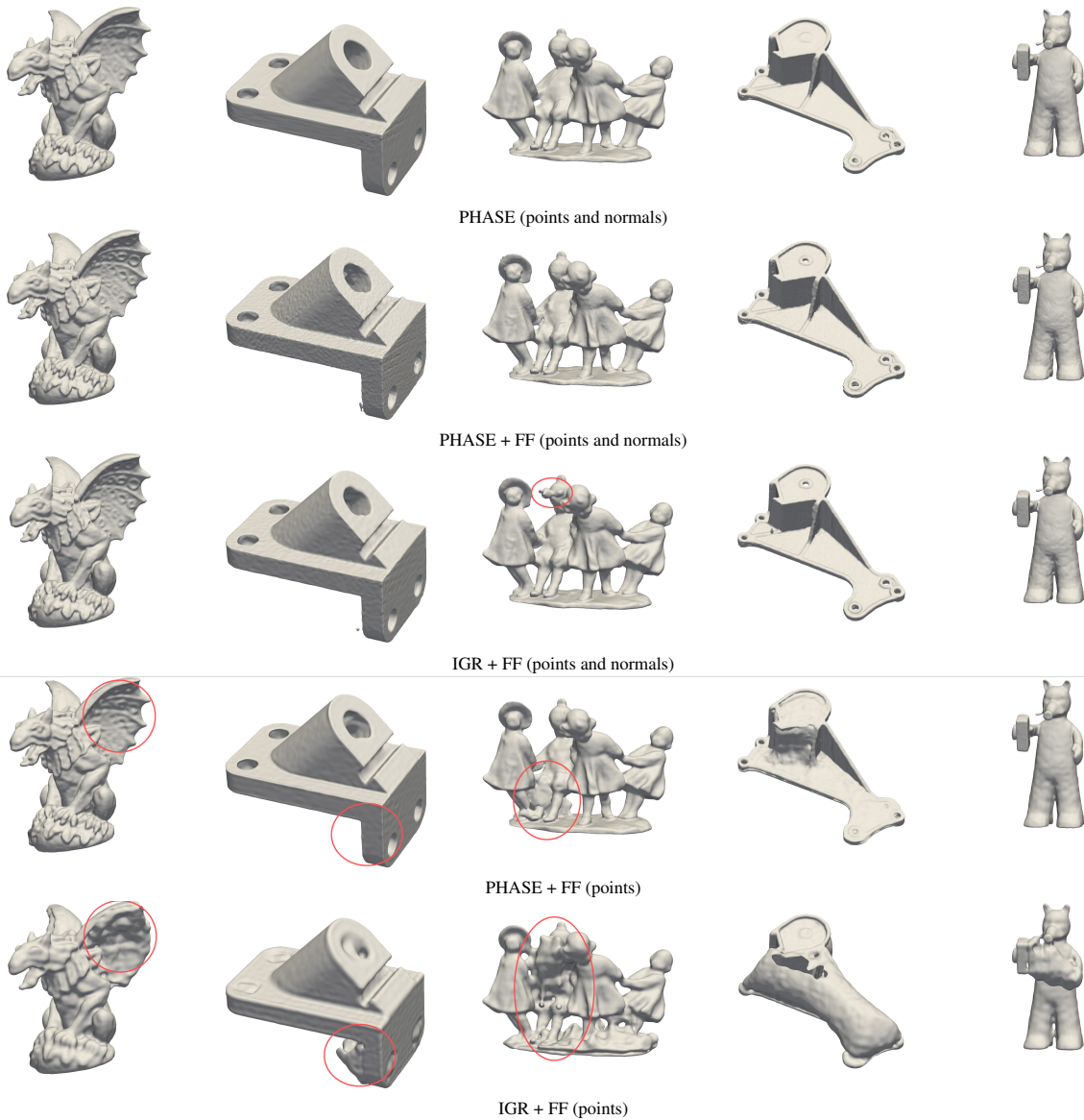


Figure 8. Surface reconstruction benchmark (Williams et al., 2019) comparing PHASE and IGR with/without normal data and Fourier features. Note that IGR has a slight tendency to add redundant surface area that is amplified in the more challenging points-only case with Fourier features, where PHASE, due to its minimal surface perimeter property is less susceptible to high frequency artifacts.

**Phase Transitions, Distance Functions, and Implicit Neural Representations**

---

Model	Method	Ground Truth		Scans	
		$d_C$	$d_H$	$d_C^{\rightarrow}$	$d_H^{\rightarrow}$
Anchor	PHASE+FF	0.24	5.30	0.09	1.17
	PHASE	<b>0.21</b>	<b>4.29</b>	0.09	1.23
Daratech	PHASE+FF	<b>0.18</b>	<b>2.53</b>	0.08	1.79
	PHASE	<b>0.18</b>	2.92	0.08	1.80
DC	PHASE+FF	0.15	2.32	0.05	2.77
	PHASE	0.15	2.52	0.05	2.78
Gargoyle	PHASE+FF	<b>0.16</b>	3.68	0.06	0.87
	PHASE	<b>0.16</b>	<b>3.14</b>	0.07	1.09
Lord Quas	PHASE+FF	0.12	<b>0.84</b>	0.04	0.94
	PHASE	<b>0.11</b>	0.96	0.04	0.96

Table 3. PHASE+FF results on the surface reconstruction benchmark of (Williams et al., 2019). See also Table 1 in main paper for more methods.

parameter search and chose the one that provides lowest error metrics on the Gargoyle model. Figure 8 shows the results of both. Note that IGR is much more sensitive to the high frequency bias and adds extraneous surface parts. In contrast, PHASE, due to its minimal perimeter property is much less susceptible to high frequency bias, although is not completely immune to this bias for point cloud data alone, as can be inspected in the image. For point cloud and normal data, we found PHASE to be even more robust to the high frequency bias, see PHASE+FF (points and normals) experiment above.